Preliminary Master Use Case List (12/15/24)

CSE416, S01 – Fall 2024

This draft of the use case list is meant to help you think about requirements. You should think about additional use cases that serve the project goals and also think about more detail for each use case.

Project goals

Perform economic and demographic analysis of political data, specifically:

- Analyze the distribution of racial, ethnic, and economic groups within states;
- Determine if district boundaries in states disadvantage certain groups;
- Analyze in-state voting behavior of racial, ethnic and economic groups;
- Compare results of one state with a high average household income with another state with a low average household income; and
- Identify Congressional district plan patterns along economic lines.

Terminology

- Population – the population of a region (e.g., state, precinct, etc.) refers to the total population as defined by the US Census Bureau. Some calculations might refer to the voting age population (VAP) or citizen voting age population (CVAP). You can use any measure of population, but it should be done consistently throughout the application.
- Ideal district population is the total population of the state divided by the number of districts in the state.
- Ensemble – a collection of district plans generated on the SeaWulf. Each such district plan will be random and will be a subset of all of the possible graph partitions that are constrained by the limits on population equality.
- Poverty level – Use the Federal Poverty Level (FPL) when analyzing relative household income levels. For aggregation purposes, you can assume a family of 5 when determining the FPL. Refer to https://www.healthcare.gov/glossary/federal-poverty-level-fpl/.
- Many of the use cases call for distribution of average household income. For these use cases, you may substitute the distribution of population in each income bucket. You can also replace this with median household income.

Notation

Use cases listed below include a categorization following the use case title. Three categories of use cases are provided as "required," "preferred," and "optional." Use cases with an "SD," "AD," or "part of GUI SD" indicate that the use case might be requested in the design review with "SD" referring to a sequence diagram and "AD" referring to an activity diagram. For some use cases, the GUI and the server part can be combined into one use case diagram. Many of the use cases that are related and that require activity diagrams can and should be combined into one activity diagram.

Some of the use cases are closely related and might be combined into a single design document. The reason for closely related use cases is that we use our use cases as units of work, so there is an attempt to provide additional use cases for some more difficult tasks.

Ensemble sizes

For testing purposes, you will need to create different size ensembles. Your maximum size ensemble should include 5,000 plans for each state.

General GUI (15 required)

GUI-1. Select state to display (required) (SD)

The user can pick a state through a dropdown menu or possibly through clicking on the state in a map of the US. The state selection will cause a table to be displayed that contains a summary of the ensembles available for the state. Ensemble data in the table will include the number of district plans in the ensemble and the population equality threshold used in the MCMC computation. In addition, state selection will also cause the map of the state to be displayed as described in GUI-2.

GUI-2. Display the current district plan when state is selected (required) (SD)

After selecting a state either from the map or the dropdown, by default, the user should be shown the current (2022) Congressional district plan displayed on the centered state map at a zoom level appropriate to the size and location of the state.

GUI-3. State data summary (required) (SD)

The data associated with the state will be summarized in response to the user selecting the state and shown concurrently with the map of the state (GUI-2). At a minimum, the summary data will include state population (either total or voting age population), state voter distribution (estimate based on a recent statewide election), population of each significant racial/ethnic group in the state, percentage of population by type of region (rural, suburban, and urban), party control of the redistricting process (if any), distribution of the population by average household income. and summary of Congressional representatives by party. You may substitute population percentage for population.

GUI-4. Display demographic heat map by precinct (required) (SD)

When the user selects a minority group from a drop-down menu, a heat map for the demographic group in the state will be displayed. The monochromatic heat map will show the percentage range of the selected group in each precinct. Choose a number of bins that effectively shows the population distribution with bin ranges that are equal. Use bounds that are integer values of population percentage. The map will include a legend that displays the bin ranges and associated colors. To improve readability, you can eliminate any bins that contain no values thereby improving the color separation.

GUI-5. Display economic heat map by precinct (required)

When the user selects an income level heat map, a monochromatic heat map will show the average household income in each precinct. Choose a number of bins that effectively shows the population distribution with bin ranges that are equal. Use bounds that are integer values of population percentage. The map will include a legend that displays the bin ranges and associated colors. A similar heat map will show the region type (rural, urban, suburban). Either heatmap can be selected from a drop down menu that also includes the option to select a demographic heat map.

GUI-6. Display poverty level heat map by precinct (required)

When the user selects a poverty level heat map, a monochromatic heat map will show, for each precinct, the percentage of population below the poverty line. Choose a number of bins that effectively shows the population distribution with bin ranges that are equal. Use bounds that are integer values of population percentage. The map will include a legend that displays the bin ranges and associated colors. This heat map can be selected from a drop-down menu that also includes the heat maps described above.

GUI-7. Display a political/income level heat map (required)

When the user selects a suitable GUI component, a heat map will show, for each precinct, a bi-chromatic heat map with a blue range of colors for Democratic precincts and red for Republican precincts. The gradations of color will be based on the household income in that precinct and the hue based on the party receiving the most votes in a selected previous election. Choose a number of bins that effectively shows the population distribution with bin ranges that are equal. Use bounds that are integer values. The map will include a legend that displays the bin ranges and associated colors. This heat map can be selected from a drop down menu that also includes the heat maps described above.

GUI-8. Display state Congressional representation table (required) (SD)

When the user clicks on screen component selecting district detail (or some other appropriate trigger), a table will be displayed. Each row in the table will contain data for one Congressional district. At a minimum, the data will contain the district number, the representative (for the enacted plan only), the representative's party, ~~the representative's racial/ethnic group~~, average household income, percentage of population below the poverty level, percentage of population by region type (rural, urban, suburban), and the vote margin as a percentage in the selected recent election.

GUI-9. ~~Display~~ Highlight district ~~plan~~ (required)

If a user clicks on some identifier of a district in the Congressional detail table, the district will be highlighted on the map. Highlighting can be performed in a variety of ways. For example, the border of the highlighted district might change color or thickness.

GUI-10. Reset page (preferred)

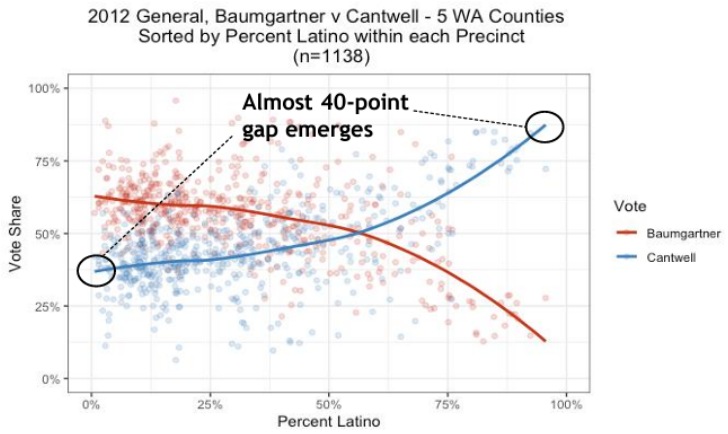
When the user clicks a reset button, the GUI will reset to the condition before the user selected a state.

GUI-11. Compare two district plans on the map (preferred) (SD)

Compare two district plans by showing both plans on the map. This could be limited to comparing a selected random plan (i.e., interesting) with the enacted plan. The trigger will be some GUI component (e.g., "Compare with enacted" button).

GUI-12. Display Gingles 2/3 analysis results (required) (SD)

In response to a user request, display a scatter plot (example below) that shows the precinct election results for each party organized on an x, y axis by percentage or racial/ethnic group in the precinct and party vote share. Any of the feasible racial/ethnic groups in the state should be capable of being displayed. The color of a dot in the scatter plot will correspond to the party receiving the most votes in the selected recent election. You might also show two dots for any precinct, a blue dot for the Democratic vote share and a red dot for the Republican vote share.



GUI-13. Display the Gingles 2/3 analysis result for precinct income (required)

In response to a user request, display a scatter plot (example below) that shows the precinct election results for each party organized on an x, y axis by average household income in the precinct (x-axis) and party vote share (y-axis). The color of a dot in the scatter plot will correspond to the party receiving the most votes in the selected recent election. The plot will also include a dropdown option to display results based on region type of each precinct (rural, urban, suburban). For the region type display, one suggestion is to divide the x-axis into three stages, 1/3 for rural, 2/3 for suburban, and 3/3 for suburban. However, you can use a different approach if it shows the results better.

GUI-14. Display the Gingles 2/3 analysis result for precinct income/race (required)

In response to a user request, display a scatter plot (example below) that shows the precinct election results for each party organized on an x, y axis by average household income / race in the precinct (x-axis) and party vote share (y-axis). The color of a dot in the scatter plot will correspond to the party receiving the most votes in the selected recent election. Values in the x-axis will be calculated as a combination (e.g., linear) of race and income in a way that provides equal consideration of each category.

GUI-15. Display the Gingles 2/3 analysis data in a tabular display (preferred)

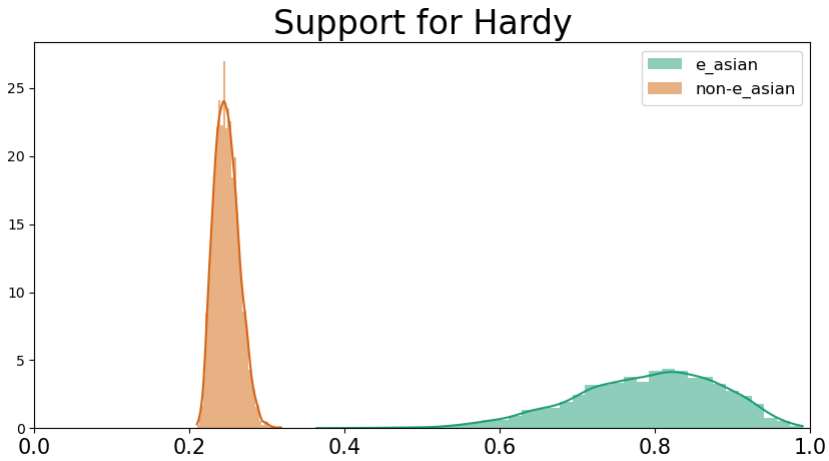
For all of the Gingles 2/3 analysis data, a table display of the precinct-by-precinct results will be displayed. Each row will display the data for a precinct including, total population, ~~population by~~ region type (rural, urban, suburban), ~~minority~~ non-white population, average household income, Republican votes, and Democratic votes.

GUI-16. Highlight a Gingles 2/3 table row (preferred)

In response to a user selecting a dot in each of the Gingles scatter plots, the precinct identified by the dot will be highlighted in the corresponding table.

GUI-17. Display candidate results of Ecological Inference (EI) analysis (required) (SD)

Display the results of the EI analysis in response to a user GUI request. The user shall have the ability to select the racial/economic/~~region~~ groups to compare. The results will be shown in a display for each candidate (example below) in which the x-axis represents the percentage of a racial/economic/~~region~~ group in the state that voted for a candidate and the y-axis represents the associated probability value for each x-axis value. ~~The racial groups will be contained in a separate chart.~~ You will combine the precinct data for race and economics in a single expression in which each term is weighted near equally. You can determine the best economic categories (e.g., below poverty, low income, high income). Optionally, you can provide an option to show racial groups and economic groups separately.



GUI-18. Display EI analysis by region type (required)

The above EI chart will have an option to calculate and display the statewide results (i.e., one statewide election) by region type (rural, urban, suburban).

GUI-19. Display EI precinct results in a bar chart (preferred)

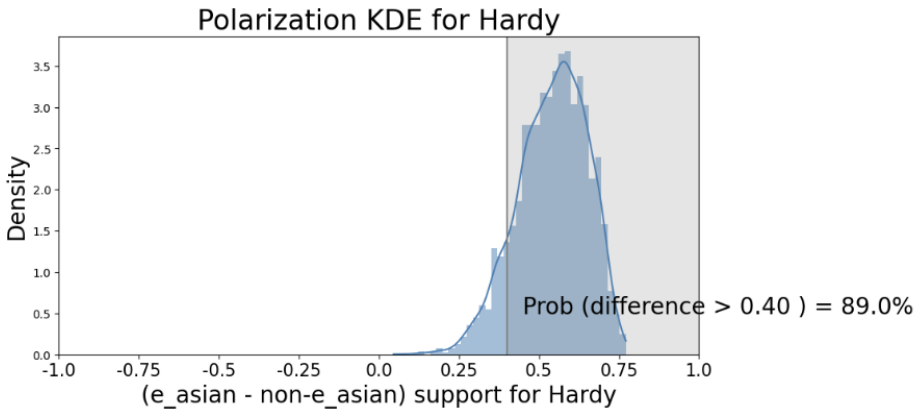
Display the EI results in a bar chart for the categories mentioned in GUI-17. The height of each bar will correspond to the peak value in the chart for each category. Each bar will also display a confidence interval showing the range of values determined in the EI analysis.

GUI-20. Display EI precinct results in a choropleth maps (preferred)

Display the EI results in a choropleth maps (one map per candidate) in which each precinct is displayed in a color that is consistent with the most likely level of support for the candidate in the precinct.

GUI-21. Display EI KDE results (preferred)

Display the EI KDE results that compare support for a candidate between two racial/ethnic groups (e.g., white and African American).



GUI-22. Display EI KDE results for racial/ethnic and income (preferred)

Display the EI KDE results that compare support for a candidate between two racial/ethnic groups (e.g., white and African American) combined with income level. The display should clearly show differences (if any) between low income white and low income African American.

GUI-23. Display box & whisker data (required) (SD)

The user will be able to request the display of box & whisker data for an ensemble of district plans and a particular racial/ethnic group from among the feasible groups in the state. There will also be an option to display results organized by region type (rural, urban, suburban) The display will show the results of the ensemble box & whisker analysis for each group. Dots for each district in the current enacted district plan will be shown in the display (in order of increasing percentage of the ~~minority group or region type~~ associated display. The display should be sufficient in size to show your largest state and should include a legend and color selection to make the chart easily readable.

GUI-24. Display vote share vs seat share curve (preferred)

If your Gingles-2/3 test shows racially polarized voting, display the vote share vs seat share curve for the state. If the Gingles-2/3 test does not indicate racially polarized voting in one of your states, the GUI component that allows the user to select the display should be disabled.

GUI-25. Display box & whisker data for income (required)

Modify the box & whisker display option so that the user can select either a racial box & whisker or an economic Box & whisker. Enacted plan should be included in the display.

GUI-26. Display box & whisker for region type (optional)

Modify the box & whisker display option so that the user can also select a regional box & whisker as an option. Regions available would be rural, urban, and suburban. Enacted plan should be included in the display.

GUI-27. Display an "interesting" district plan (preferred)

Display on the map one of the interesting plans identified in SeaWulf-6.

GUI-28. State summary - additional data (optional)

The data associated with the state will be summarized in response to the user selecting the state and shown concurrently with the map of the state (GUI-2). In addition to the data specified in GUI-3, the following summary data is also displayed: population density, median income, poverty rate, political lean of state, number of districts, and number of precincts.

GUI-29. District data summary (optional)

When a user clicks on a district on the district map, summary data for the selected district will be shown concurrently with the district map. At a minimum, the summary data will include the district voter distribution, population of each significant racial/ethnic group in the district, and distribution of the population by average household income. You may substitute population percentage for population.

GUI-30. Precinct data summary (optional)

When a user clicks on a precinct on the precinct map, summary data for the selected precinct will be shown concurrently with the precinct map. At a minimum, the summary data will include the precinct voter distribution, population of each significant racial/ethnic group in the precinct, and distribution of the population by average household income. You may substitute population percentage for population.

GUI-31. Display electoral heat map by precinct (optional)

When a user selects an electoral level heat map, a bi-chromatic heat map will show the winning party as well as the vote margin for each precinct. Democratic precincts will be shown in blue, while Republican precincts will be shown in red. The gradations of color will be based on the vote margin in that precinct, with a darker color representing a large vote margin and a lighter color representing a tight vote margin. Choose thresholds for each bin to effectively show the vote margin variation across precincts. This heatmap can be selected from a drop down menu that also includes the heatmaps described in GUI-4 through GUI-7.

Preprocessing (12 required)

Prepro-1. Integrate multiple data sources (required) (AD)

Integrate and merge US Census data (income, region type (rural, urban, suburban) population, both for total and for any opportunity groups), precinct data (boundary, name, demographics, etc.), and existing district data (boundary, name, district#, etc.). Geographic boundary data should be converted (if necessary) to a consistent format (e.g., GeoJSON).

Prepro-2. Identify precinct neighbors (required) (AD)

Identify two precincts as neighbors if they share a common boundary of at least 200 feet and the edges of each precinct are within 200 feet of its neighbors' edges. If possible, try to locate a data source for which this computation is already done.

Prepro-3. Integrate enacted plan with dataset (required)

Integrate the enacted plan for the state within the server database.

Prepro-4. Store preprocessed data (required)

The preprocessed data should be stored in the NoSQL or relational database. If a relational database is used, the data should be stored in third normal form. Data might also be stored in a file system accessible to the server.

Prepro-5. Store SeaWulf data (required)

Retrieve generated data from SeaWulf for each of your states, convert to an appropriate format, and store either in your database or in a file system. Data stored in a file system should be accessible through a path obtained from your database.

Prepro-6. Calculate statewide measures (required)

Overall state measures include percentage of Republican voters, percentage of Democratic voters, percentage of each of the demographic groups relevant for the state and percentage of population by region type (rural, urban, suburban). A demographic group is considered relevant for a state if the population of the group is sufficient to provide a winning vote margin in at least one district. You should estimate the percentage of voters in each party by analyzing the votes for each party in each precinct in some statewide election (e.g., presidential, attorney general, governor).

Prepro-7. Generate data files required for SeaWulf processing (required) (AD)

Generate all the data files required for SeaWulf processing. This will include the graph representation of the precincts in a state as well as geographic, election, and incumbent data for each precinct.

Prepro-8. Gingles 2/3 precinct analysis (required) (AD)

Perform a precinct-by-precinct analysis of voting results and minority population percentage for some statewide race (2022 or 2020). For each precinct, the analysis will identify the winning party, the Republican vote share, the Democratic vote share, the population percentage of each significant racial/ethnic group, the population by region type (rural, urban, suburban), and the average household income. The analysis is repeated for each feasible racial/ethnic group in the state.

Prepro-9. Gingles 2/3 non-linear regression analysis (required) (AD)

For the statewide race used in the use case above, calculate the non-linear regression curve for the Republican and Democratic precinct values for each Gingles 2/3 graph. Multiple equation forms will be used to determine the best form for non-linear regression.

Prepro-10. Use the PyEI MGGG software to calculate Ecological Inference data (required)

Use the PyEI MGGG software to calculate results for ~~multiple statewide races~~ a selected statewide race (e.g., presidential, gubernatorial, etc.).

Prepro-11. Calculate the vote share vs seat share curve data (preferred)

Using the Shen software as a starting point, calculate the data for the vote share vs. seat share curve in any of your states that display racially polarized voting. Use the current district plan as the basis for the calculation. Also use relatively fine grain increments of vote share and possibly randomization to reduce a stair-stepping effect.

Prepro-12. Calculate region type for each precinct (required)

Calculate the region type (rural, urban, and suburban) for each precinct. You can use some reasonable values for the ratio of area to population to determine the classification of each precinct.

Prepro-13. Calculate Box & Whisker Data for Enacted Plan (required)

Calculate the box & whisker data for the enacted district plan for any of the box & whisker displays attempted.

SeaWulf (8 required)

SeaWulf-1. Server dispatcher (required)

Establish a protected directory on SeaWulf to store your team's data. Pre-stage any data that might be used repeatedly for SeaWulf runs. Prior to submitting a batch districting run request to SeaWulf, the data required for the run should be marshalled (from memory and/or DB) and passed to the SeaWulf as a file (or multiple files) to be stored in the team's SeaWulf file system.

SeaWulf-2. Run MGGG ReCom algorithm on the SeaWulf (required) (AD)

Set the constants in the MGGG code to define the properties (e.g., constraints) of the run. Any run-control information should be packaged in a SeaWulf acceptable format (e.g., script commands) and executed on SeaWulf. Your activity diagram should demonstrate that you understand how the MGGG algorithm operates. You should generate a test ensemble and a large ensemble. The test ensemble will contain approximately 250 random district plans and the large ensemble will contain approximately 5,000 plans.

SeaWulf-3. Coordinate/aggregate SeaWulf core generated data (required) (AD)

You will run your code on a single SeaWulf node, one that has multiple cores. Each of the cores will generate one or more random graph partitions (i.e., district plans) and store a concise version of those results in a shared file directory for your team. Following the completion of each random graph partition, the core begins the generation of the next random district plan. You will coordinate the work of the multiple cores so that when the target number of district plans is completed, each of the cores ends its processing.

SeaWulf-4. Calculate election winners (required) (AD)

Using 2022 statewide election results, estimate the election results in each district of each ensemble district plan. You will calculate this by summing up the estimated votes in each node (i.e., precinct) of a partition sub-graph. You can use a suitable precinct by precinct vote in a 2022 statewide election. If there is no 2022 statewide election, you can use the 2020 presidential election results.

SeaWulf-5. Calculate the Republican/Democratic split for each random district plan (required) (AD)

For each generated plan in an ensemble, estimate the Republican/Democratic votes in each district. Since each district is a collection of precincts, use the historic precinct vote totals (e.g., 2020 Presidential, 2022 Gubernatorial, etc.) to estimate the winner of an election in each district.

SeaWulf-6. Identify and store additional random district plans of note (~~required~~ preferred) (AD)

You will not be able to store all your random district plans in the server database, but you will store some subset of those plans. Summary information and detailed information of such plans should be stored in your server database for eventual display by the user. A district plan of note includes the plans with many districts that are heavily rural, urban, suburban and ones with districts with interesting economic patterns (e.g., those with a minimum and maximum deviation from the statewide average household income). Teams should decide what is "interesting." About 5-10 plans would be sufficient. The "interesting" plans would be displayed in GUI-11.

SeaWulf-7. Calculate ensemble measures (required) (AD)

Calculate the summary measures for each ensemble. At a minimum, measures will include the number of district plans, ~~deviation from statewide average household income~~, and Republican/Democratic splits.

SeaWulf-8. Run on multiple SeaWulf nodes (preferred)

Run the MGGG algorithm and supplemental code on multiple nodes on SeaWulf. Node coordination should be done using a suitable coordination mechanism (e.g., MPI). Be able to estimate speed-up during your final project presentation. Note that this is optional since your algorithms should run effectively on a large-core single node processor.

SeaWulf-9. Python profiler (preferred)

Profile your system performance on SeaWulf using a Python profiler tool. Identify the procedures that consume the most CPU time. Results can be displayed using some Python-appropriate tool and displayed as an image in your final presentation.

SeaWulf-10. Removed

This use case is kept to maintain numbering of use cases. The contents were merged with SeaWulf-6.

SeaWulf-11. Calculate box & whisker data (required) (AD)

Calculate the box & whisker summary data for all the random district plans generated by the SeaWulf. These calculations will made be for each feasible racial/ethnic group in the state.

SeaWulf-12. Calculate box & whisker data (required) (AD)

Calculate the box & whisker summary data for all the random district plans generated by the SeaWulf. These calculations will be made for each economic group in the state. This data will be displayed in GUI-25.

SeaWulf-13. SeaWulf-1. Calculate box & whisker data (preferred) (AD)

Calculate the box & whisker summary data for all the random district plans generated by the SeaWulf. These calculations will be made for each region group in the state. This data will be displayed in GUI-26.

Server Processing (0 required)

All server processing addressed in GUI use cases.

Additional Use Cases